

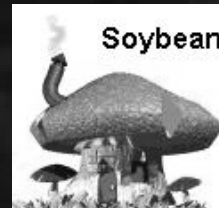


Nematodes, blueberries, and chickens; Applied bioinformatics at the USDA

Benjamin F. Matthews

USDA-ARS

Soybean Genomics & Improvement Laboratory



Major topics

- Bioinformatics applied to agricultural problems studied in my laboratory
 - Types of data
 - Bioinformatics applications
- Other agricultural problems at USDA-Beltsville

Projects

- Soybean interactions with cyst nematode
- Soybean interactions with rust
- Soybean mapping –SNP discovery
- Blueberry cold hardiness
- Cocoa fingerprinting
- Chicken interaction with coccidiosis
- Swine interaction with nematode (Ascaris)
- Spiroplasma kunkellii genome sequencing project

Problem:

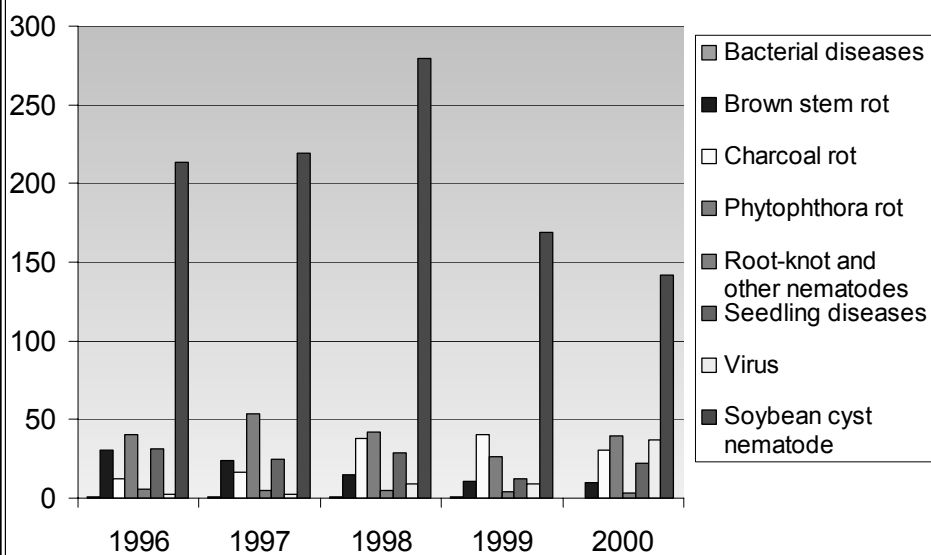
Soybean cyst nematode damage

- Heterodera glycines
- Affects most growing regions in the US
- Found world-wide
- 7% of the crop in the US lost
- Approx. \$1 billion lost each year in US
- Multiple genotypes of SCN

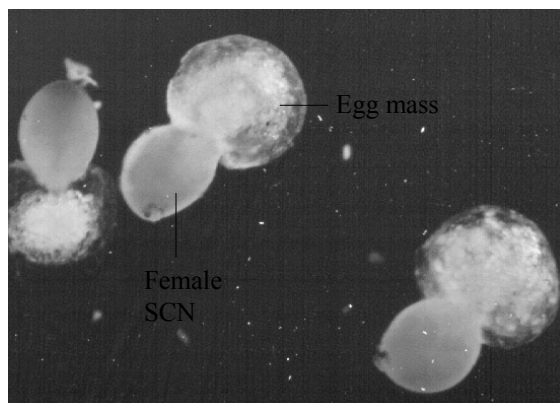
Soybean cyst nematode damage



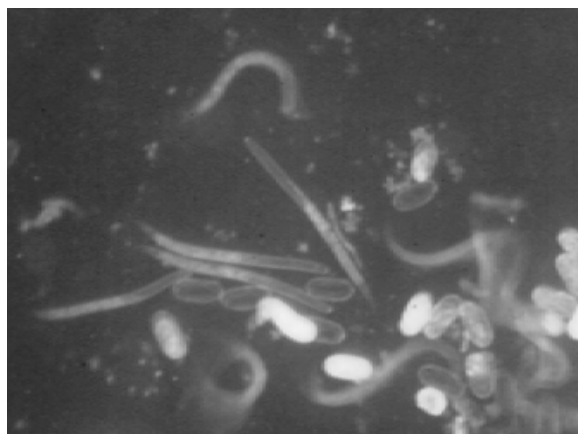
Soybean Losses (millions of bushels)



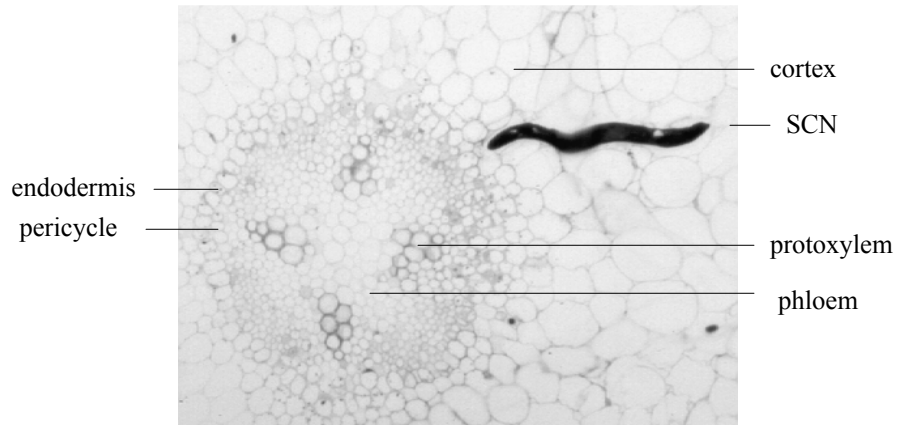
Extruded egg masses



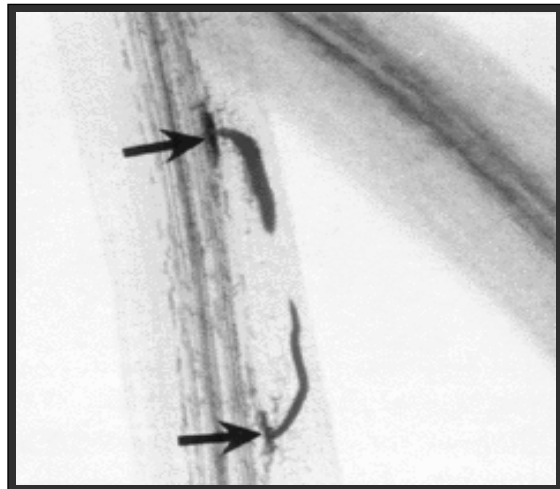
Nematodes hatching



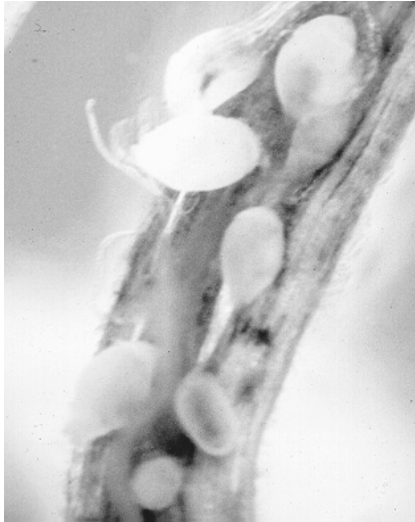
Nematode invading root



SCN establishing syncytium



SCN female



Solution: Broaden Resistance of Soybean to SCN

- Two Approaches
 - I. Clone known resistance genes
 - Map position-based cloning
 - Move cloned genes into elite varieties
 - II. Develop new modes of resistance
 - Determine mechanisms conferring resistance
 - Gene expression profiles, metabolic profiling, cytology
 - Identify pathways important to resistance
 - Design new resistance?

I. Clone known resistance genes

- Gene Mapping
 - Mapmaker; Joinmap
 - Mapmaker QTL; QTL Cartographer;
 - Single marker ANOVA-SAS
 - Place molecular markers on map
 - Place resistance loci (phenotype) on map
- Use markers close to resistance to obtain continuous DNA clones
 - LaserGene Seqman
 - (FPC) v 4.7.9 (Fingerprinted Contigs)
- Obtain DNA Sequence
- Identify genes in that sequence
 - GeneMine
- Test candidate resistance genes

Map-based isolation

Genetic maps (Location of phenotypes via genetic crosses)



**Physical map of genome (actual DNA sequence)
or portion of genome**



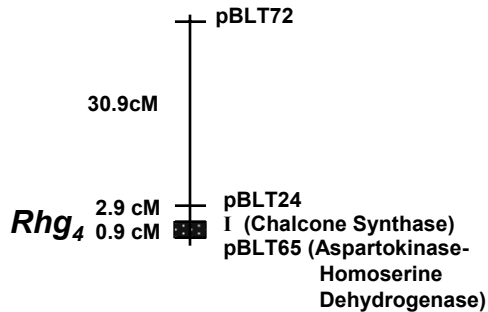
Identify genes in the DNA sequence



**Annotate DNA sequence
Location of promoter elements
Introns, exons, enhancers, etc.**

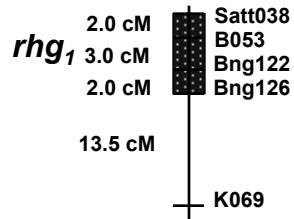
SCN resistance linkage groups

LG A2



Population:
BARC-2 x Noir 1
149 F₂ individuals

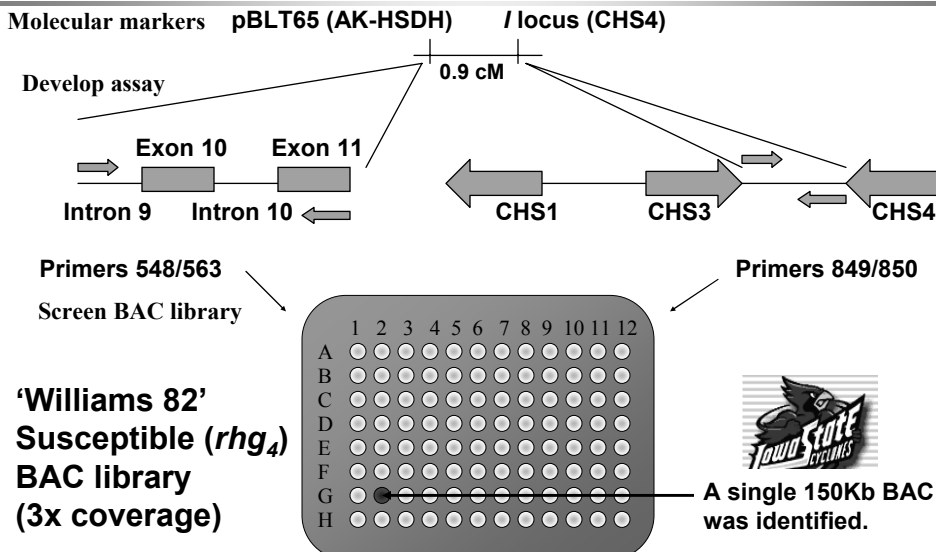
LG G



Population:
G. max x *G. soja*
56 F₂ individuals

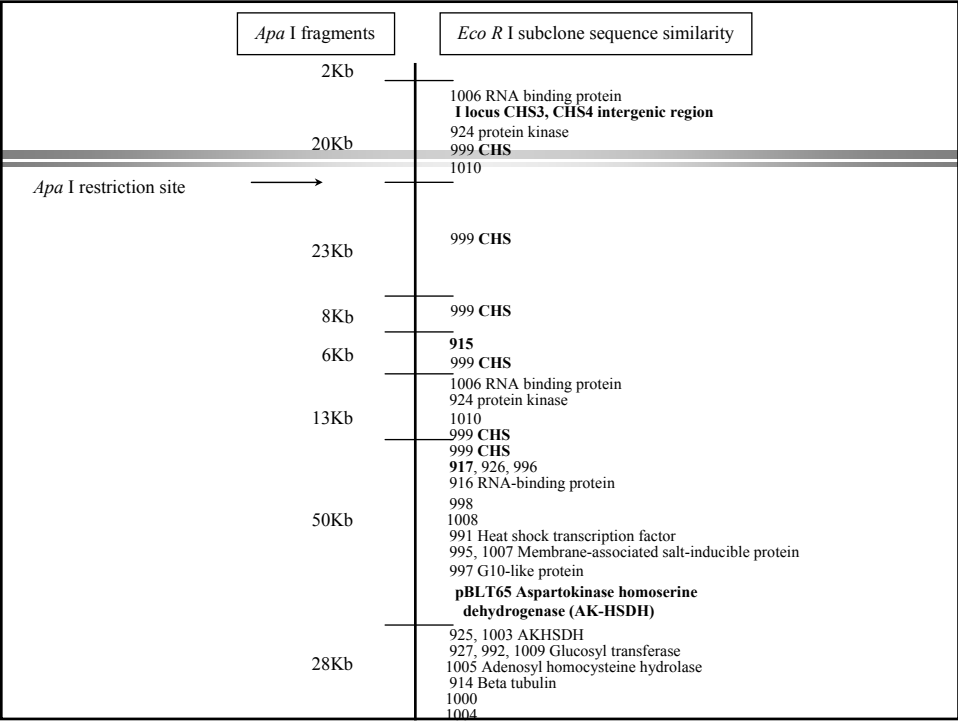
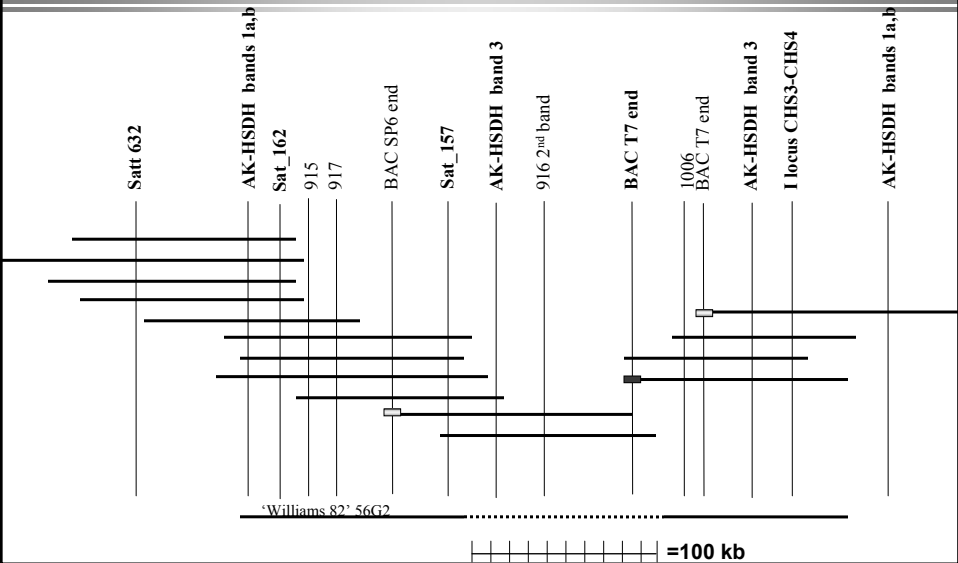
From a genetic map to a physical map

Genetic Linkage Group A2





PI 435.654 contig vs. Williams 82 BAC



Summary

- Genetic mapping
 - Mapmaker, JoinMap, QTL Cartographer
- Identify BAC clones
- Align BAC clones
 - FPC, LaserGene SeqMan
- DNA sequence
 - Trim vector, align, continuous sequence
- Annotate sequence –
 - GeneMine
 - Komal Kaul, UMUC

II. Develop new modes of resistance

What genes and pathways are important?

- Gene profiling
 - Microarrays
 - RT-PCR
- Metabolic profiling
- Cytology
- Identify target genes
- Provide tools
 - Promoters
 - DNA sequences of target genes

Soybean cDNA libraries

- Soybean leaves and roots two days post SCN invasion
- Subtractive 10 hr
- 12 hr root
- 2 & 4 day
- 6 & 8 day
- Differential display
- Cotyledon 96 clones

Soybean Genomics & Microarray Database (SGMD)

- EST database
Expressed Sequence Tag – One-pass 5' DNA sequence
- Database to track clones
- Organize clone information
 - Location
 - DNA sequence
 - Identity using BLAST
 - BLAST score
 - Expression levels

<http://bldg6.arsusda.gov/benlab/>

EST data import

- DNA sequences are cleaned to remove vector and contaminating sequences
- EST sequences are batch BLAST compared to GenBank sequences
- BLAST results are extracted and imported into SGMD

Portion of data from plate A05

A05 Clones

Clone ID	dbEST_ID	GenBank_Accn	Name (Best Blast hit)	Clone Type	E-value	Score	Identities	% Identities	Blast N	Blast X
A05A01	10346069	BM107938	ubiquitin conjugating protein-like [Arabidopsis thaliana]	cDNA	2.2E-49	524	100/161	62	A05A01BlastN	A05A01BlastX
A05A02	10346070	BM107939	putative nitrate transporter NRT1-3 [Glycine max]	cDNA	4.2E-71	729	151/179	84	A05A02BlastN	A05A02BlastX
A05A03	10346071	BM107940	dihydroflavonol 4-reductase-like [Arabidopsis thaliana]	cDNA	2.6E-60	501	93/135	68	A05A03BlastN	A05A03BlastX
A05A04	10346072	BM107941	unknown protein [Arabidopsis thaliana]	cDNA	0.000000000026	171	36/48	75	A05A04BlastN	A05A04BlastX
A05A05	10346073	BM107942	cellobiohydrolase I [Trichoderma viride]	cDNA	2.1E-67	694	122/180	67	A05A05BlastN	A05A05BlastX
A05A06	10346074	BM107943	hypothetical protein F17M5.260 - Arabidopsis thaliana	cDNA	5E-34	326	62/136	45	A05A06BlastN	A05A06BlastX
A05A07	10346075	BM107944	cytochrome P450 [Arabidopsis thaliana]	cDNA	8.9E-22	255	45/105	42	A05A07BlastN	A05A07BlastX
A05A08	10346076	BM107945	hypothetical protein T14P8.16 - Arabidopsis thaliana	cDNA	2.3E-73	635	120/151	79	A05A08BlastN	A05A08BlastX

<http://bldg6.arsusda.gov/benlab/>

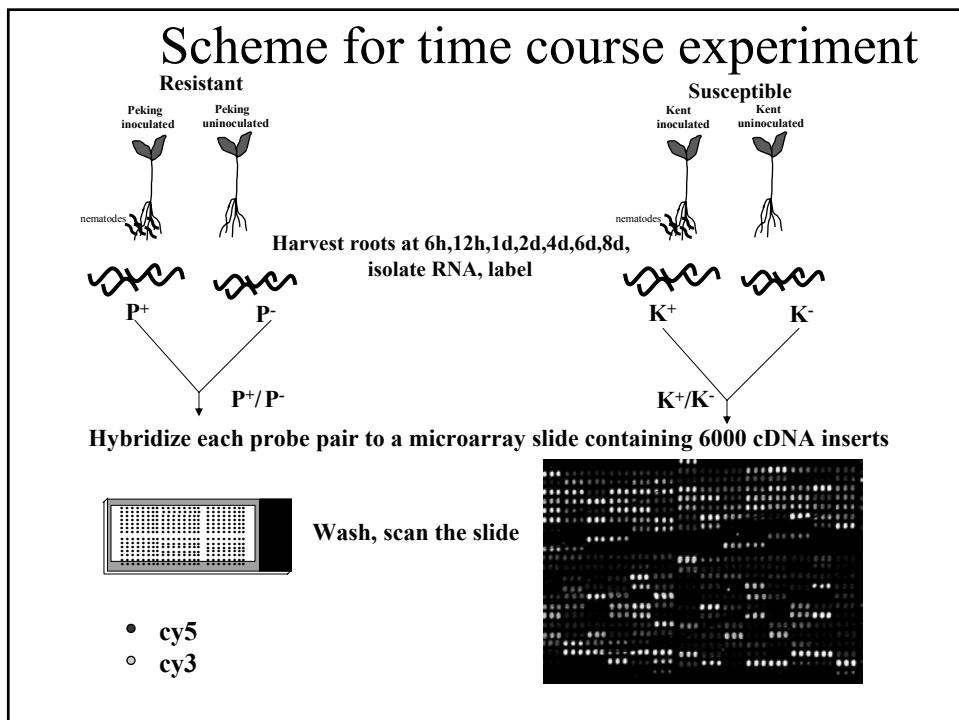
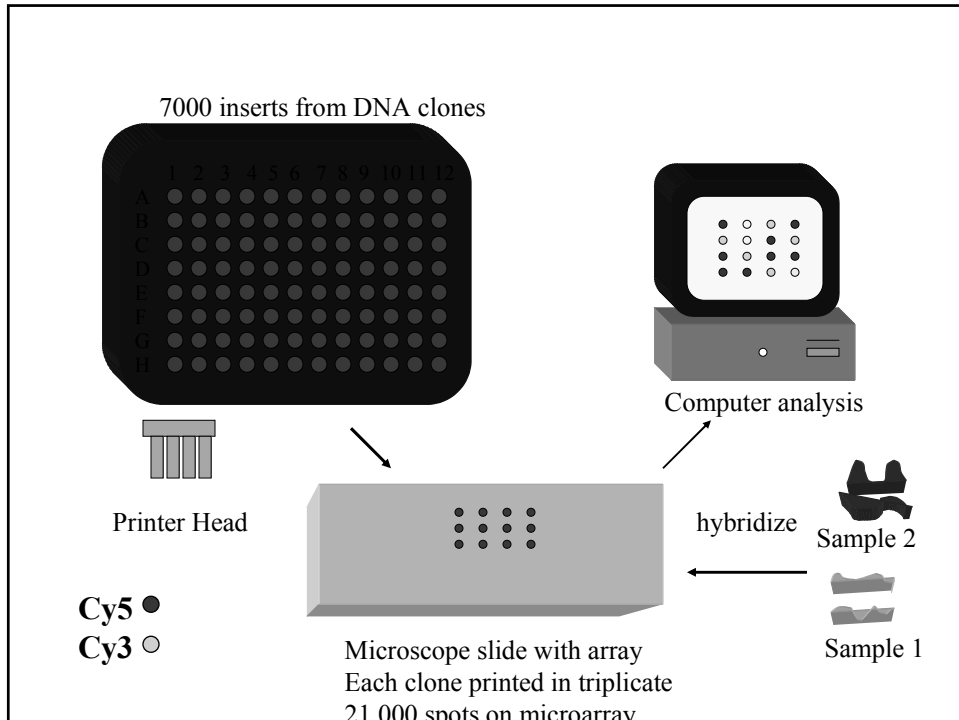
Soybean Genomics and Microarray Database

- Established in 1999
- EST and microarray database
- Approximately 8,000 soybean and 4,000 nematode ESTs
- Over 5 million rows of microarray data
- Built on SqlServer2000, but SQL code is universal
- MIAME compliant

<http://bldg6.arsusda.gov/benlab/>

SGMD

- First soybean microarray database that is publicly available
- Integrates analysis with design
 - No need to resort to third party software
 - Binary queries easy to perform
 - Online advanced queries
- Allows users to download data as html, text or XML formats



Microarray data analysis

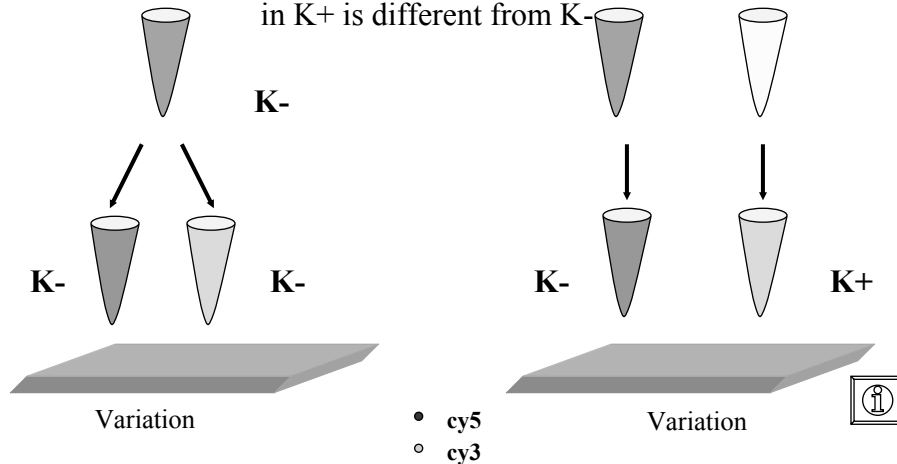
- Slides are scanned
- Raw data from all images are extracted using SPOT
- Data are imported into database
- Data are filtered to eliminate flagged spots and very low intensity spots
- Data are normalized using Lowess
- Reproducibility is determined by ANOVA

- Online analytical processing (OLAP) tools are used
- Significance of induction is determined using T-tests
- Scripts written in SQL and integrated into the database

T-test K^-/K^- vs K^+/K^-

Kent = susceptible
Peking = resistant

What is the probability that
the expression level of a gene
in K^+ is different from K^-

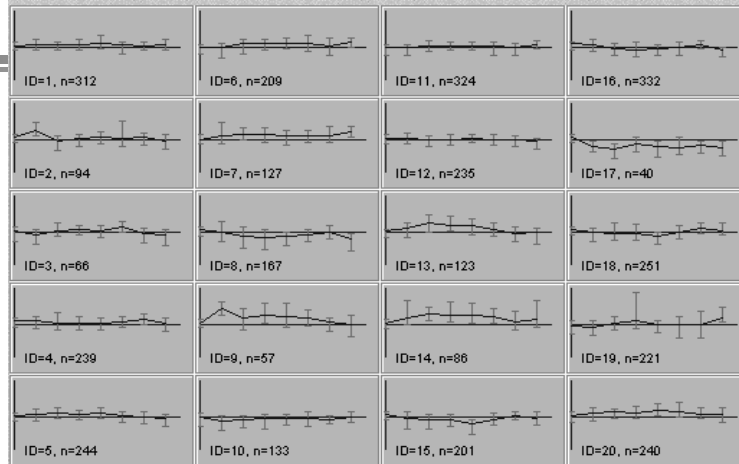


Microarray data Self Organizing Maps

- Group genes according to similarity of expression level over time
- Number of groups can be controlled



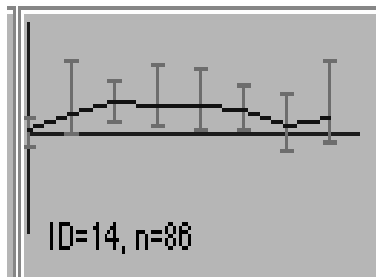
K-means clustering of P+/P- time series data:



A large number of genes exhibit no change in expression
Genes are induced or suppressed across several time points

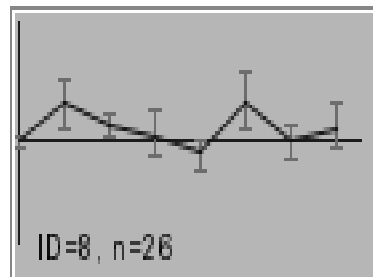
Distribution of genes in the two clusters into different functional categories

P+/P- cluster 14



Signaling pathways 5
Defense 8
Metabolism 15
Cell wall formation/maintenance 12
Unknown 46

K+/K- cluster 8



Signaling pathways 1
Defense 6
Metabolism 10
Cell wall formation/maintenance 4
Unknown 19

P+/K-



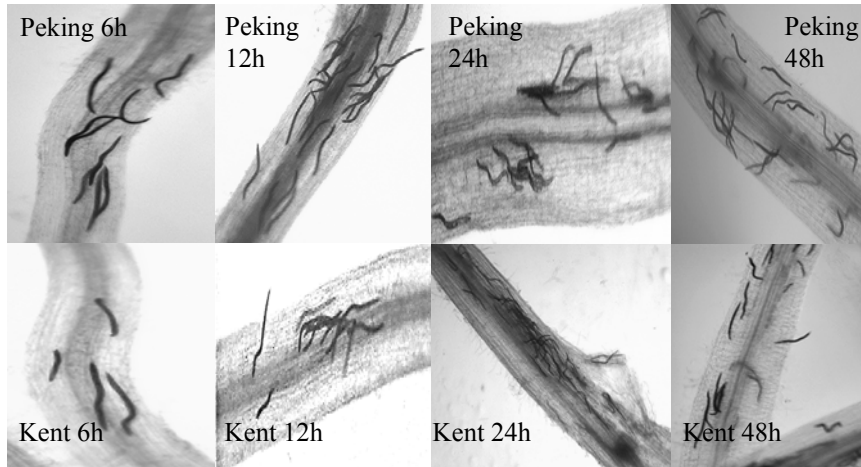
Note: $\text{LOG}_2(5) = 2.32192809488736$

Clone ID	Putative Name	Cy3/Cy5 Log Ratios	E-Value	Blast N	Blast X
A01A03	<i>chalcone synthetase, chs7</i>	4.039187104	3.8e-17	A01A03BlastN	A01A03BlastX
A01A04	<i>BAC T19G15, from chromosome V near 60.5 cM</i>	4.231761475	2.4e-25	A01A04BlastN	A01A04BlastX
A01A18	<i>leucine zipper protein HAT14, homebcx (p46665)</i>	2.886116153	9.1e-08	A01A18BlastN	A01A18BlastX
A01A20	<i>protein kinase, putative</i>	3.54280888	2.1e-52	A01A20BlastN	A01A20BlastX
A01A22	<i>SBT1- subtilisin-like protease</i>	7.596830686	1.7e-34	A01A22BlastN	A01A22BlastX
A01B02	<i>genomic DNA, chromosome 5, TAC clone: K19B1</i>	9.576087286	9e-09	A01B02BlastN	A01B02BlastX
A01B06	<i>histone deacetylase, putative</i>	3.481484531	0.022	A01B06BlastN	A01B06BlastX
A01B13	<i>peroxidase, cationic (gi577503)</i>	3.442509721	2.6e-53	A01B13BlastN	A01B13BlastX
A01C02	<i>lymphocyte-activation gene 3 (LAG3) mRNA</i>	4.19174034	0.55	A01C02BlastN	A01C02BlastX
A01C03	<i>putative host response protein (Pir7)</i>	5.492684841	1.7e-10	A01C03BlastN	A01C03BlastX
A01C08	<i>4-coumarate:CoA ligase</i>	3.012707473	1.7e-50	A01C08BlastN	A01C08BlastX
A01C11	<i>chromosome II BAC F19D11 genomic sequence</i>	3.742174394	0.12	A01C11BlastN	A01C11BlastX
A01C13	<i>poymorpha chloroplast genome DNA</i>	6.94303649	0.094	A01C13BlastN	A01C13BlastX
A01C15	<i>galactosidase, beta (e1285876)</i>	4.425387157	3e-27	A01C15BlastN	A01C15BlastX
A01C16	<i>hypothetical protein, phosphotyrosine, (Q02191)</i>	3.04298235	2.1e-07	A01C16BlastN	A01C16BlastX
A01C18	<i>Soybean mRNA for reductase involved in deoxychalcone synthesis (NAD(P)H dependent 6'-deoxychalcone synthase)</i>	3.588471727	3e-71	A01C18BlastN	A01C18BlastX

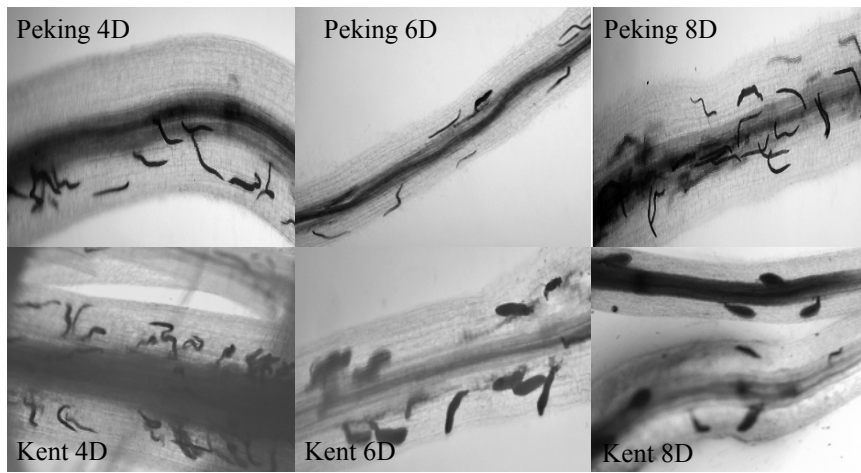
Cytological data

- What is happening at cellular level
- Microscopic examination
- In situ hybridization

Timeline of SCN at 6, 12, 24 , 48 h after infection in a resistant(Peking) and susceptible (Kent) soybean cultivar



Timeline of SCN at 4, 6, 8 days after infection in a resistant(Peking) and susceptible (Kent) soybean cultivar



Soybean Cellular Response to SCN

18 hrs
ER and organelles
increase
Cell walls
dissolve

3 to 4 days
Cells deteriorate
Cell walls thicken
Cell wall deposition
ER accumulation

Syncytium
nearly empty
Necrosis of
cells

Resistant

6 to 24 hrs 24 to 48 hrs Syncytium collapses
4 to 5 days

Nematode penetration and migration Syncytium established

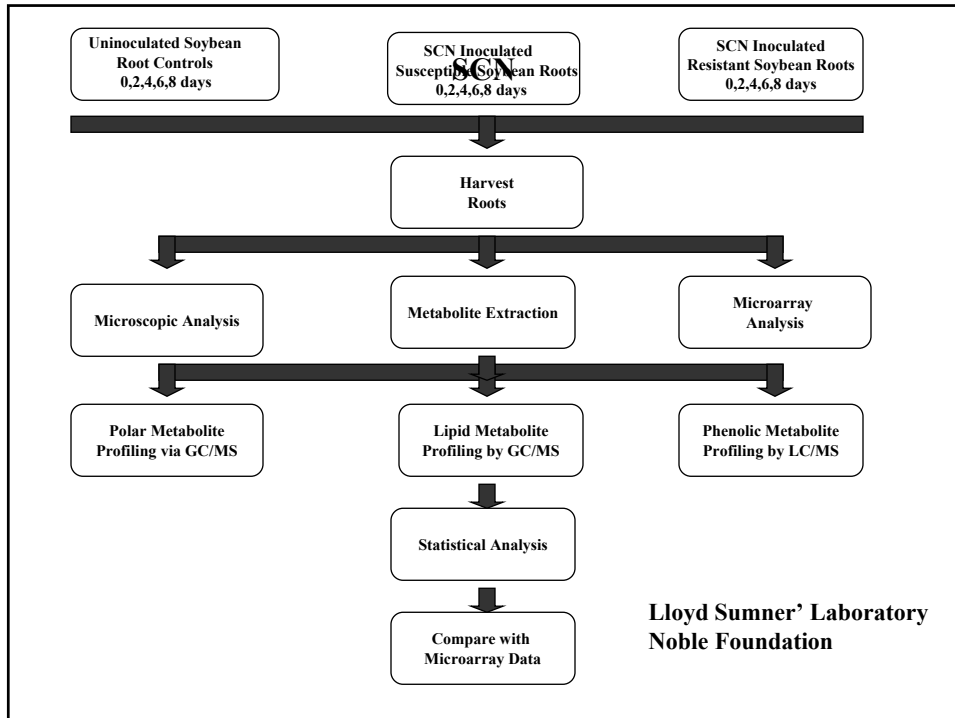
Susceptible

18 hrs
Dense cytoplasm
Enlarged vacuole
Increase in ribosomes and
rough Endoplasmic Reticulum

3 to 4 days
Cell wall dissolution
Thickened cell wall
Membrane proliferations
Accumulation of ER, mt, granules, vacuoles
Callose deposition
Finger-like walls near xylem

Metabolic profiling data

- Measure a range of metabolites
 - 600-1000 different metabolites
 - Identify each metabolite
- Perturb system
 - Eg. (+) and (-) nematodes – genes induced
- Correlate gene expression profile with metabolite profile and cellular events



Changes in metabolites

- Sucrose increases
- Trehalose increases
- Long chain fatty acids increase (Kent)
- Valine levels differ (P vs K)
- Unknown secondary compound
 - present in P absent in K

Biological significance of data

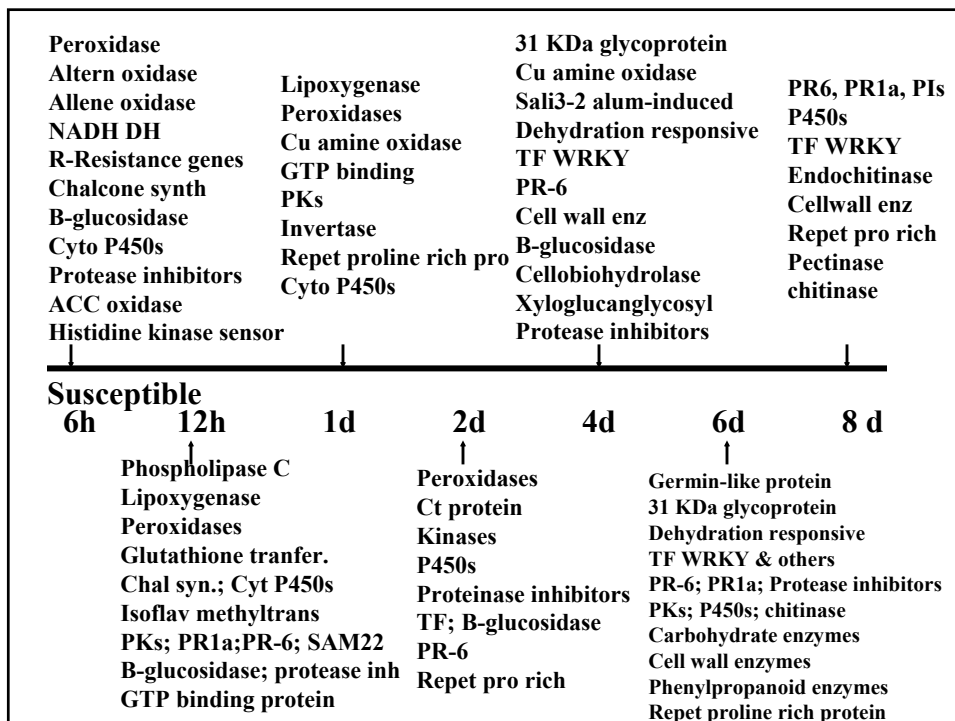
- Integrate gene expression profiles, cytological data, metabolic profiles
- Metabolic pathways
- Signaling pathways
- Interactions among proteins & pathways
- Literature - known facts
- Develop new insights and hypotheses
- New knowledge

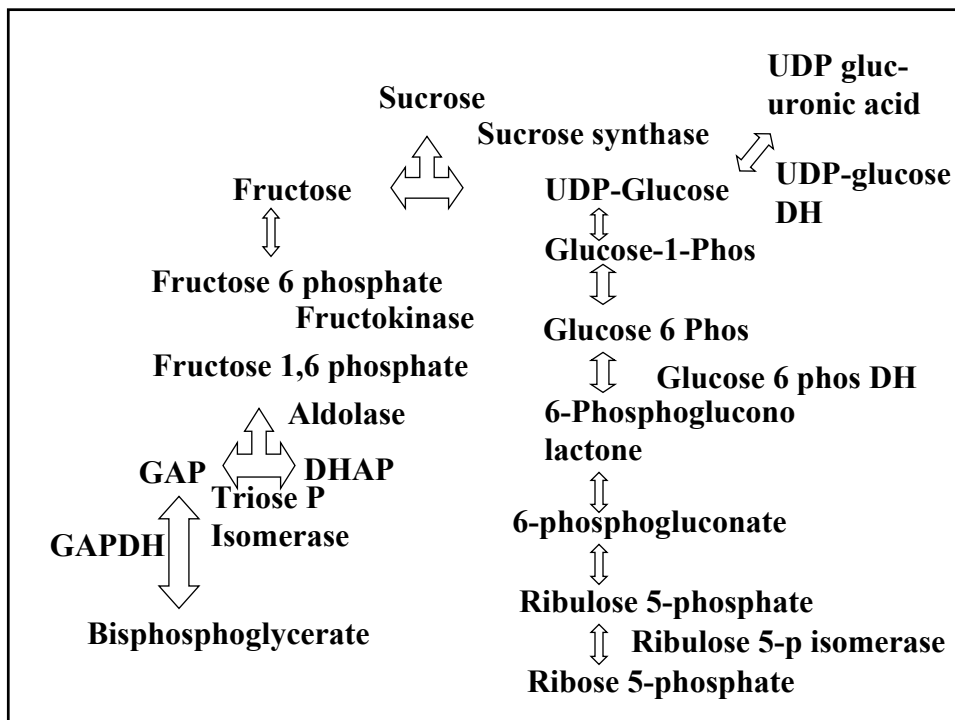
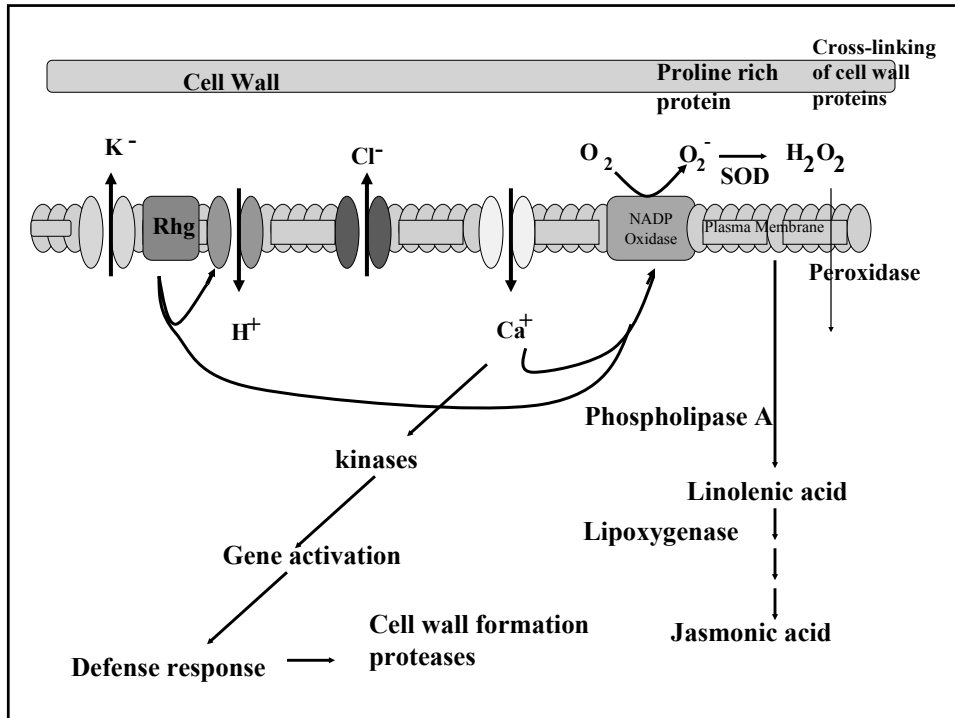
Elucidate the biological significance of the data

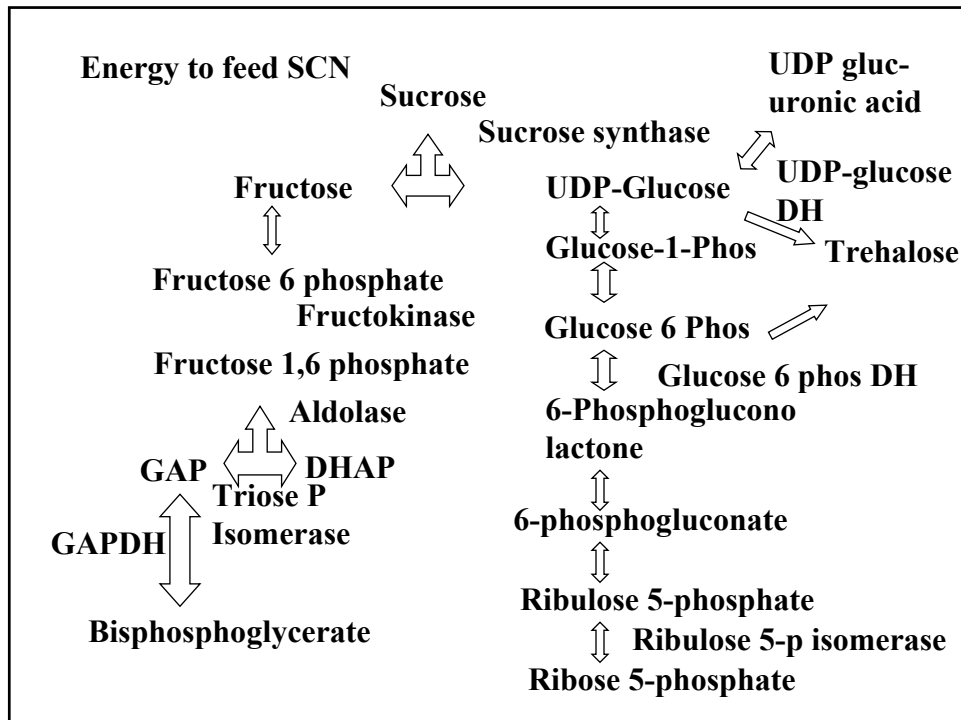
- See relationships
 - Organize data
 - Query data
 - Visualize data
- Identify key components
- Develop new hypotheses
- Data are:
 - multidimensional
 - from very different sources
 - very different types

Data

- Gene expression data
 - Microarray, EST data; northern; RT-PCR
- Some gene relationships are known
 - Families, pathways, interactions
- Cellular relationships
 - Timing; changes in organelles & structures
- Metabolic relationships
- Genetic relationships
 - Phenotypes
 - Knockouts
 - Gene silencing
- Software can predict, wet lab can test





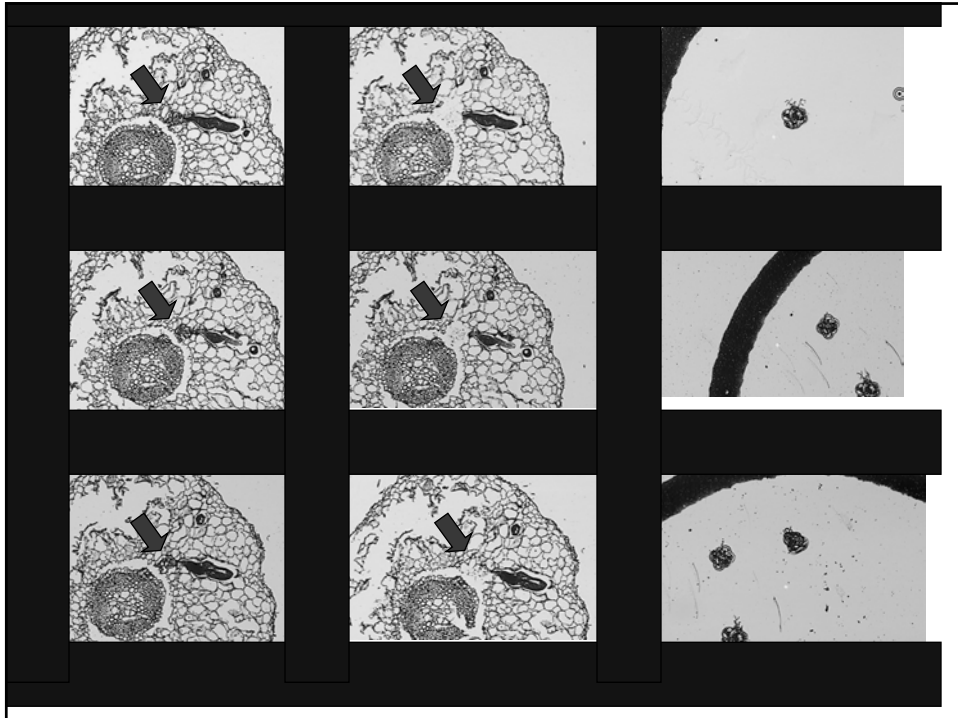


Laser Capture Microdissection

- collect a homogeneous population of syncytial cells – 7, 14, 21 dai
- Determine gene expression patterns in those cells

Arcturus PixCell IIe
LCM System



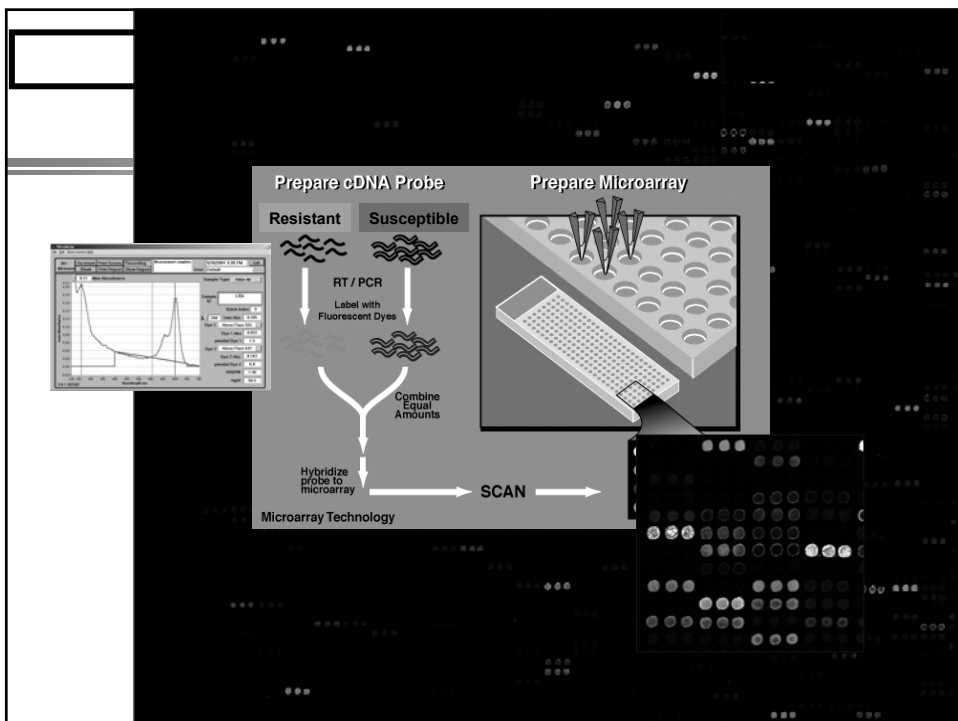
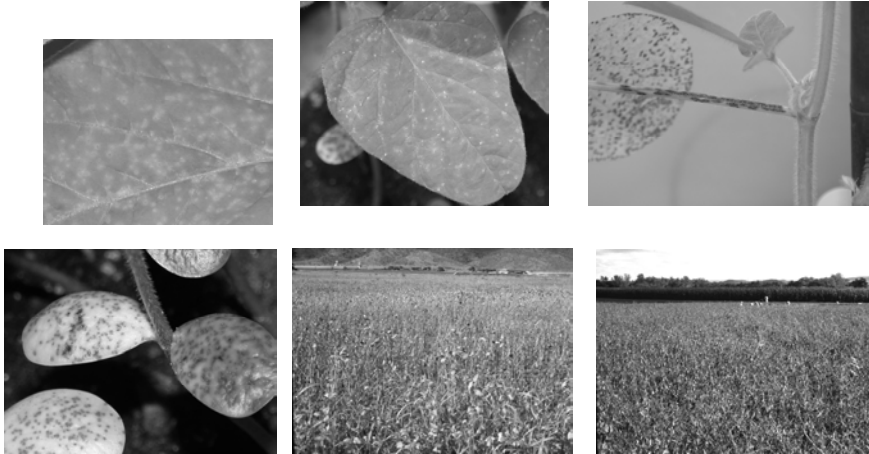


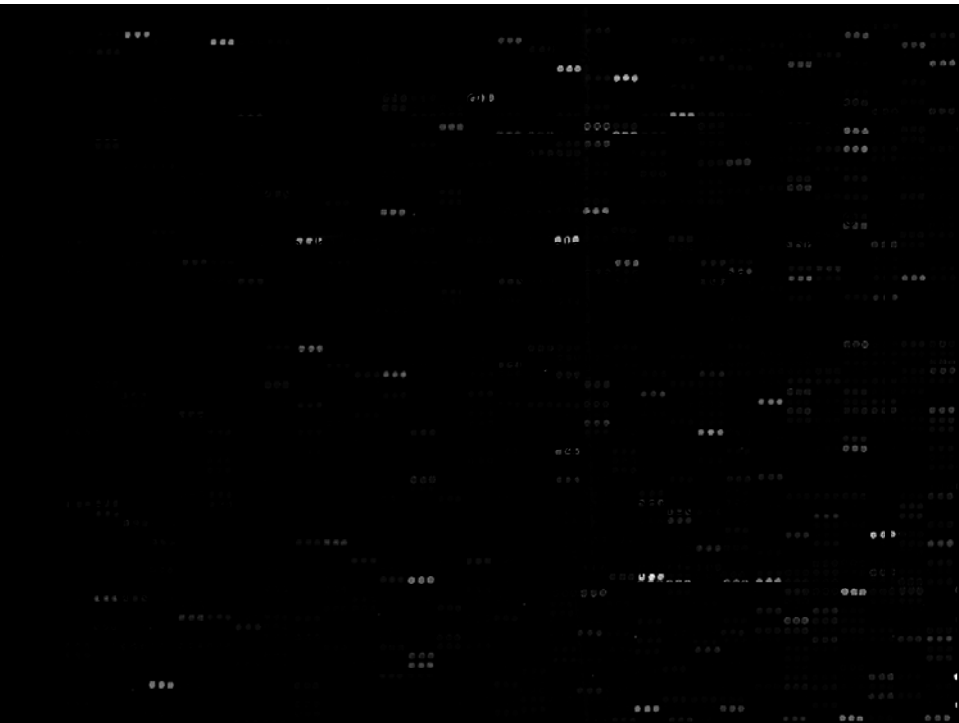
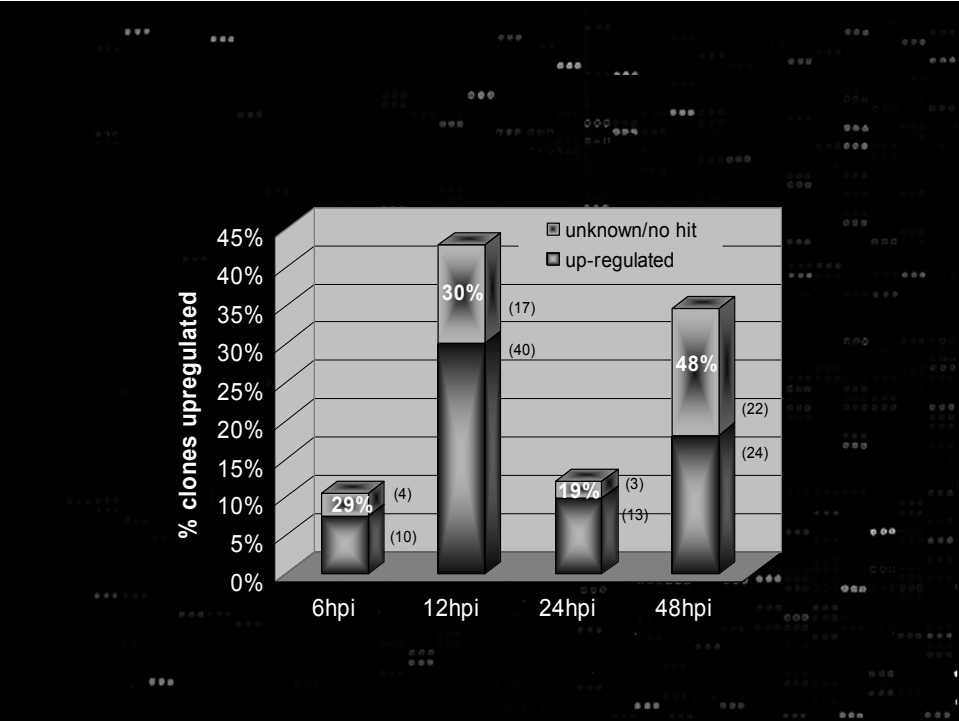
Soybean Rust

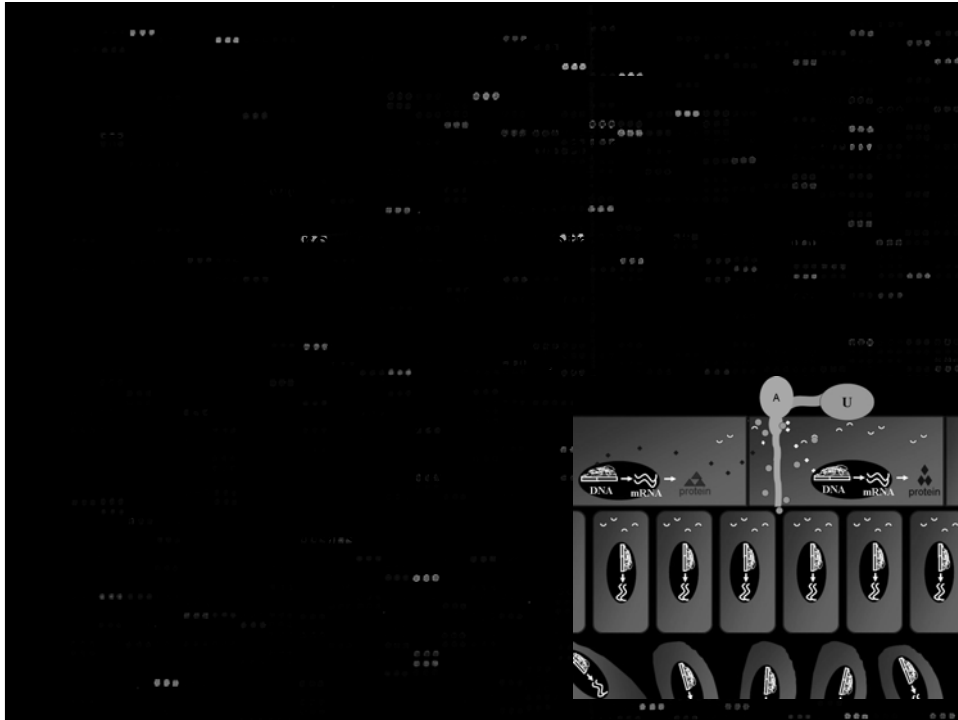
Ben Matthews, Jane Choi,
and Reid Frederick
USDA-ARS

Introduction

Symptoms on Host (soybean)

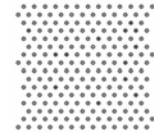
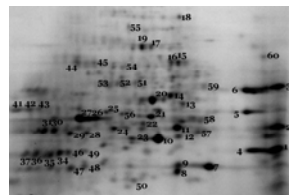






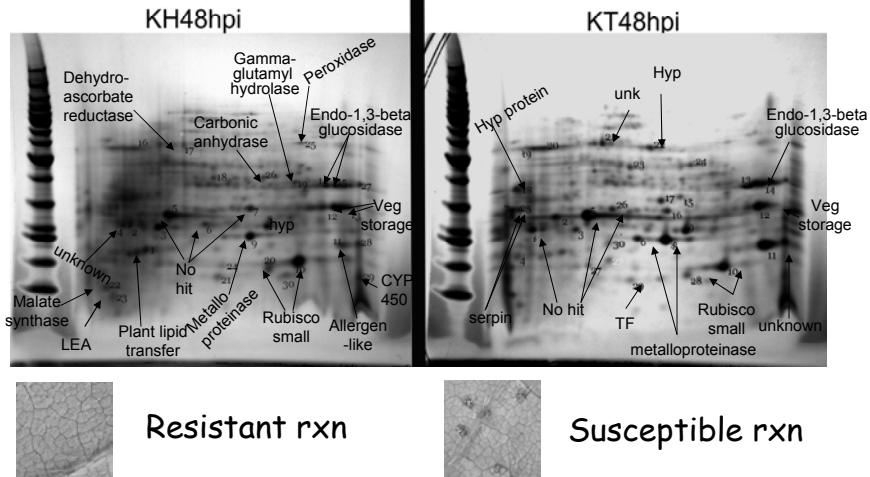
Enriched extracellular proteins from soybean leaves from resistant and susceptible interaction

- 2-D protein gel
- Pick spots for MALDI
- In-gel trypsin-digestion
- MALDI/TOF-TOF
mass spectrometry
ABI4700

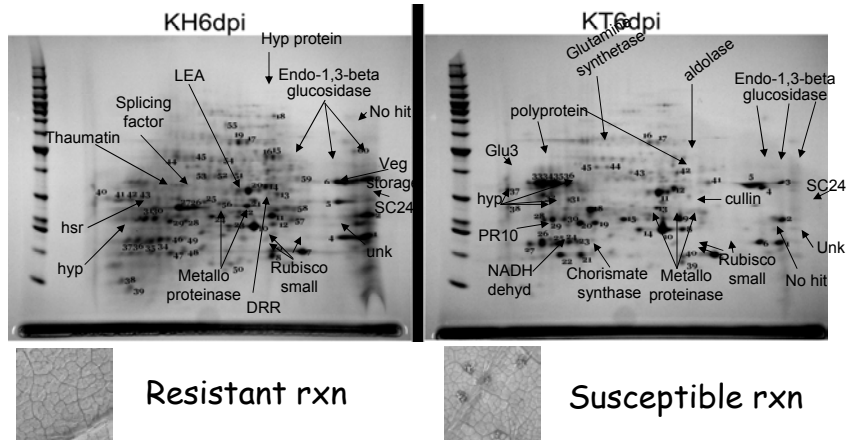


(**Jane Choi** in collaboration with Mike McMahon/Doug Luster &
Alberto Nunez, Core Technologies, USDA-ARS-ERRC)

Protein profiling with 2D-gel and MALDI/TOF-TOF mass spec 48hpi



Protein profiling with 2D-gel and MALDI/TOF-TOF mass spec 6dpi



Summary

- Microarrays show differential gene expression
- Protein gels identify protein differences
- Provides targets for modification to improve soybean resistance to rust

Soybean mapping

- DNA sequencing
- Align DNA sequences
- Identify single nucleotide polymorphisms
- Soybean, strawberry, & cattle projects

- Dr. Perry Cregan (301) 504-5070
 - Lakshmi Matukumalli, GMU student

Blueberries

- Cold tolerance-survival during winter
 - What genes provide protection
 - Blueberry ESTs
 - Microarrays
-
- Dr. Jeannie Rowland (301) 504-6654

Strawberries

- Develop a strawberry map
 - SSR mapping
 - Automate search of GenBank for SSRs
 - Primer design
 - Map position
-
- Dr. Kim Lewers (301) 504-6768

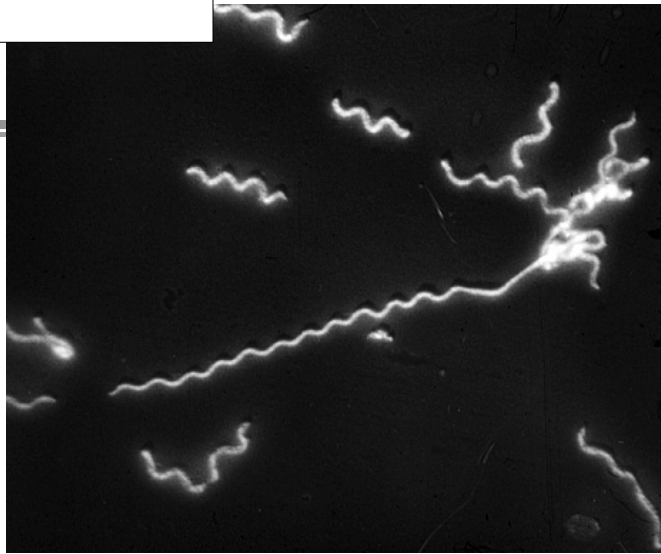
Cocoa fingerprinting

- Develop a database to deposit and query cocoa fingerprinting data
- Interface with the International Cocoa Germplasm Database
- Characteristics of 20,000 cocoa accessions

- Dr. Dapeng Zhang (301) 504-7477

Spiroplasma Kunkellii Genome sequencing Project

- Pathogen – eg. corn stunt
- Small genome
- Sequence nearly complete
- Preliminary annotation complete
- Need gap closure
- Compare *S. kunkellii* genome sequence with other genomes
- Software available-some configuration and modification required
- Dr. Bob Davis (301) 504-7545



Spiroplasma in artificial culture

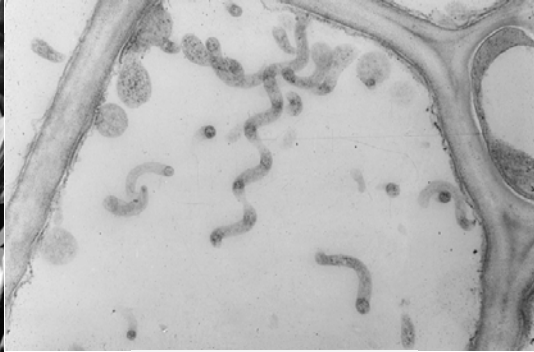
(dark field microscope view)



Leafhopper, insect vector that carries the spiroplasma



Infected corn plant



Spiroplasma, inside
a corn plant cell

Chicken-coccidiosis

- Disease of poultry
- What genes provide protection?
- Apply to poultry breeding
- ESTs
- Microarrays
- Dr. Hyun Lillehoj (301) 504-6170

Swine from nematodes

- Eliminate nematodes
 - Microarrays
 - ESTs
 - Gene homologs
-
- Dr. Joe Urban (301) 504-8765
 - Dr. Harry Dawson (301) 504-9412